BAYESIAN ANALYSES OF NONHOMOGENEOUS AUTOREGRESSIVE PROCESSES

TECHNICAL REPORT NO. 18

T. W. ANDERSON, NOZER D. SINGPURWALLA, AND REFIK SOYER

SEPTEMBER 1986

U. S. ARMY RESEARCH OFFICE

CONTRACT DAAG29-85-K-0239

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

BAYESIAN ANALYSES OF NONHOMOGENEOUS AUTOREGRESSIVE PROCESSES

T. W. Anderson

Stanford University

Nozer D. Singpurwalla* and Refik Soyer*

The George Washington University

TECHNICAL REPORT NO. 18

September 1986

Also issued as Technical Report GWU/IRRA/Serial TR–86/12,

Department of Operations Research, The George Washington University

U. S. Army Research Office

Contract DAAG29-85-K-0239

Theodore W. Anderson, Project Director

Department of Statistics

Stanford University

Stanford, California

Approved for Public Release; Distribution Unlimited.

* Research sponsored by the Office of Naval Research Contract N00014-85-K-0202 (NR 042-372) and by the U.S. Army Research Office Grant DAAG 29-84-K-0160.

Abstract

In this paper we consider nonhomogeneous autoregressive processes which are special cases of the vector-valued autoregressive processes considered by Anderson (1978) for the analysis of panel survey data. We point out that, for a nonhomogeneous autoregressive process of order higher than one, the least-squares estimates cannot be obtained unless repeated measurements are made on the time series. We present here two Bayesian approaches based on Kalman filter models which alleviate the above difficulty and result in an alternative strategy for the analyses of nonhomogeneous autoregressive processes. In our first approach the notion of exchangeability plays a key role, whereas for our second approach, which results in an adaptive Kalman filter model, an approximation due to Lindley facilitates the necessary computations for inference.

Key words: Nonhomogeneous autoregressive processes, random coefficient autoregressive processes, exchangeability, adaptive Kalman filtering, panel survey data, cross-section studies, Lindley's approximation

1. Introduction and Overview

To keep the introduction simple, we shall focus attention on a first-order autoregressive process of the form

$$y_{\alpha t} = \theta_{\alpha t} y_{\alpha, t-1} + u_{\alpha t}, \quad \alpha=1,\ldots,N; \; t=1,\ldots . \tag{1.1}$$

The autoregressive coefficients $\theta_{\alpha t}$ are assumed unknown and the innovations $u_{\alpha t}$ are assumed to be independent and normally distributed with a known mean and variance. When $\theta_{\alpha t} = \theta_t$ for all α, (1.1) will be referred to as a *nonhomogeneous* (or *inhomogeneous*) *autoregressive process*. When $\theta_{\alpha t} = \theta_\alpha^*$ for all t, (1.1) will be referred to as a *random coefficient autoregressive process*. The above nomenclature is in keeping with the terminology of Anderson (1978) and Liu and Tiao (1980), respectively, who have written on the above processes.

Nonhomogeneous and random coefficient autoregressive processes have a wide applicability in the analysis of economic, sociological, biological and industrial data. Such processes can be easily motivated in the context of "panel surveys," that is, surveys in which several respondents are interviewed at more than one point in time. Analyses of such data are sometimes called "cross-section studies" by econometricians. [See Hsiao (1986).] Anderson (1978) cites several examples of panel surveys in the economic, medical and sociological contexts and develops inference procedures for a set of several sequences of observations from the same nonhomogeneous vector-valued process. The approach taken by Anderson (1978) is least-squares with an accompanying asymptotic theory. Liu and Tiao (1980) address the panel survey problem via random coefficient autoregressive processes

which are stationary, that is, with $|\theta_\alpha^*|<1$, and propose a Bayesian approach for inference about the θ_α^*'s . The Bayesian set-up of Liu and Tiao (1980) assumes that the θ_α^*'s are independent drawings from a rescaled beta distribution.

In this paper, we present two Bayesian approaches for inference in a nonhomogeneous autoregressive process of order $p \geq 1$. The process considered by us is a special case of the vector-valued nonhomogeneous autoregressive processes considered by Anderson (1978). A motivation for the p-th order nonhomogeneous autoregressive process has also been given by Horigome, Singpurwalla and Soyer (1985) who consider the problem of monitoring for "reliability growth." The data from reliability growth problems can be regarded as being the result of a panel survey.

In Section 2 we introduce the vector-valued nonhomogeneous autoregressive process of Anderson (1978) and review the least squares estimators of the parameters of this process. We point out that for such processes with $p > 1$, it is not possible to obtain the least squares estimators unless N is also greater than one. We contrast this with the Bayes estimators which do not suffer from such restrictions.

The set up of Section 3 can be cast as an ordinary Kalman filter model, whereas that of Section 4 can be cast as an *adaptive* Kalman filter model. The term *adaptive filtering* is used in the engineering literature whenever some or all of the parameters of the observation or the state equation of the Kalman filter are estimated from the data [Broemeling (1985), p. 274]. A review of the different approaches to adaptive filtering is

given by Mehra (1972). With adaptive Kalman filtering, the automatic and closed form nature of the ordinary Kalman filter [cf. Meinhold and Singpurwalla (1983)] is lost. Shumway (1983) has considered maximum likelihood estimation in adaptive Kalman filtering using the *expectation-maximization* algorithm of Dempster et.al. (1977). A Bayesian approach to adaptive Kalman filtering has been considered by Magill (1965) but Magill's treatment assumes that the unknown parameters of the linear system can only take a finite number of distinct values. The approach suggested by us here does not have such a restriction and uses an approximation due to Lindley (1980) which enables us to obtain computable results. Our use of Lindley's approximation for the analysis of adaptive Kalman filter models is new and it represents a contribution, albeit a minor one, to the state of the art of filtering.

In Section 3 we present our first approach. The notion of *exchangeability* plays a key role in our development here - it enables us to assign a structure of dependence for the coefficients of a nonhomogeneous autoregressive process of order $p \geq 2$ and $N \geq 1$. Such a structure of dependence alleviates the requirement that N be greater than one.

In Section 4 we present our second approach. Here we confine our attention to the case $p = N = 1$, but assume that the coefficients of the nonhomogeneous autoregressive process are themselves described by a homogeneous autoregressive process of order one, with an unknown coefficient. Thus the structure of dependence of Section 4 is stronger than that of Section 3, but with $p = 1$, the model of Section 4 is simpler than that of Section 3.

2. Least Squares Estimation in Nonhomogeneous Autoregressive Processes

Suppose that $\underset{\sim}{y}_t$ is an m-component column vector and $\underset{\sim}{\Theta}_t$ an m x m time-variant matrix of coefficients. Let $\{\underset{\sim}{u}_t\}$ be a sequence of mutually independent m-component vectors, $\underset{\sim}{u}_t$ having a normal distribution with mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{U}_t$; the index t=1,2,..., denotes time. A first-order vector-valued nonhomogeneous autoregressive process is of the form

$$\underset{\sim}{y}_t = \underset{\sim}{\Theta}_t \underset{\sim}{y}_{t-1} + \underset{\sim}{u}_t, \quad t=1,\ldots, \tag{2.1}$$

where $\underset{\sim}{y}_0$ is assumed known.

If there are N distinct units (or individuals) in a survey, and m measurements are taken for each unit, then we will observe N different time series. Thus for example, $\underset{\sim}{y}_{\alpha t}$ is an m-component vector of measurements on the α-th individual at time t.

Given $\underset{\sim}{y}_{\alpha t}$, $\alpha=1,\ldots,N$, and $t=1,\ldots,T$, the least-squares estimator of $\underset{\sim}{\Theta}_t$, obtained by Anderson (1978) is:

$$\hat{\underset{\sim}{\Theta}}_t = \underset{\sim}{C}_t(1)\ \underset{\sim}{C}^{-1}_{t-1}(0) \tag{2.2}$$

where

$$\underset{\sim}{C}_t(j) = \frac{1}{N} \sum_{\alpha=1}^{N} \underset{\sim}{y}_{\alpha t}\ \underset{\sim}{y}'_{\alpha,t-j}. \tag{2.3}$$

and $\underset{\sim}{y}'$ denotes the transpose of a column vector $\underset{\sim}{y}$.

Note that the estimators $\hat{\underset{\sim}{\Theta}}_t$ are based on the <u>pooling of information</u> from all of the N time series.

If we extend (2.1) to the case of a p-th order nonhomogeneous autoregressive process, then

$$\underset{\sim}{y}_t = \underset{\sim}{\Theta}_{1t}\ \underset{\sim}{y}_{t-1} + \underset{\sim}{\Theta}_{2t}\underset{\sim}{y}_{t-2} + \ldots + \underset{\sim}{\Theta}_{pt}\underset{\sim}{y}_{t-p} + \underset{\sim}{u}_t, \tag{2.4}$$

with $\underset{\sim}{y}_0, \underset{\sim}{y}_{-1},\ldots,\underset{\sim}{y}_{-(p-1)}$ assumed known.

The least-squares estimators of the unknown elements of the p unknown m x m matrices are also obtained by Anderson (1978); these are

$$(\hat{\underset{\sim}{\Theta}}_{1t},\ldots,\hat{\underset{\sim}{\Theta}}_{pt}) = (\underset{\sim}{C}_t(1),\ldots,\underset{\sim}{C}_t(p)) \begin{bmatrix} \underset{\sim}{C}_t(1,1)\ldots\underset{\sim}{C}_t(1,p) \\ \vdots \quad\quad \vdots \\ \underset{\sim}{C}_t(p,1)\ldots \underset{\sim}{C}_t(p,p) \end{bmatrix}^{-1} \tag{2.5}$$

where $\underset{\sim}{C}_t(j)$ is given by (2.3) and

$$\underset{\sim}{C}_t(i,j) = \frac{1}{N} \sum_{\alpha=1}^{N} \underset{\sim}{y}_{\alpha,t-i}\, \underset{\sim}{y}'_{\alpha,t-j}.$$

For the case m=N=1, that is, when we have only one measurement per item at time t, say y_t, and only one item to observe, then $(\underset{\sim}{\Theta}_{1t},\ldots,\underset{\sim}{\Theta}_{pt})$ simplifies as $\underset{\sim}{\theta}_t$ where $\underset{\sim}{\theta}_t$ is a column vector with elements $(\theta_{1t},\ldots,\theta_{pt})$ and the equation for the least squares estimator of $\underset{\sim}{\theta}_t$ is

$$\hat{\underset{\sim}{\theta}}'_t \,(\underset{\sim}{y}^{(p)}_{t-1}\, \underset{\sim}{y}^{(p)}_{t-1}{}') = y_t\, \underset{\sim}{y}^{(p)}_{t-1}{}' \tag{2.6}$$

where

$$\underset{\sim}{y}^{(p)}_{t-1} = (y_{t-1},\ldots,y_{t-p})'.$$

Note that $(\underset{\sim}{y}^{(p)}_{t-1}\, \underset{\sim}{y}^{(p)}_{t-1}{}')$ is the *outer product* matrix, and is of rank 1. Thus when m=N=1 and p > 1, the least-squares estimators (which under this set-up are also the maximum likelihood estimators) of the coefficients of the p-th order nonhomogeneous autoregressive processes are not uniquely defined.

For $p = 1$ the least squares estimators do of course exist and these take the following simple and intuitive form

$$\hat{\theta}_t = y_t/y_{t-1}. \tag{2.7}$$

In Section 3 we shall obtain Bayes estimators of $(\theta_{1t},\ldots,\theta_{pt})$ for the case m=N=1, and show that these can always be obtained and are unique. It is important to note that in obtaining Bayes estimators we are incorporating some additional structure to the model, the nature of which will be clarified in the sequel. The additional structure compensates for the lack of information due to the limitation imposed by N being equal to one.

3. Bayesian Estimation in Nonhomogeneous Autoregressive Processes Assuming Exchangeability of Coefficients.

In this section we first consider the p-th order nonhomogeneous autoregressive process (2.4) with m=N=1 and discuss inference for $\underset{\sim}{\theta}_t$. Later on we extend our results to processes with N > 1. In some applications it may be reasonable to assume a time pattern for the $\underset{\sim}{\theta}_t$'s; see for example, Section 4. However, in most instances this may not be true and what may be reasonable is some form of dependence among the vectors $\underset{\sim}{\theta}_1,\underset{\sim}{\theta}_2,\ldots,$. A simple way of describing such dependence is to assume that the sequence of column vectors $\underset{\sim}{\theta}_t$ is *exchangeable;* that is, $\underset{\sim}{\theta}_1,\underset{\sim}{\theta}_2,\ldots.$ are invariant under permutations. Exchangeability describes a mild form of dependence and this is most easily obtained by assuming that the $\underset{\sim}{\theta}_t$'s are generated by some multivariate distribution G, indexed by a vector of *hyper-parameters* $\underset{\sim}{\lambda}$, on which a prior distribution π is assigned.

It may be of interest to note that if G is not specified but estimated from the data, then the above set up would be referred to as *empirical Bayes*, whereas if G were specified but the uncertainty about $\underset{\sim}{\lambda}$ not described by π but instead $\underset{\sim}{\lambda}$ estimated from the data, then the above set up would be referred to as *parametric empirical Bayes* [cf. Morris (1983)]. With both G and π completely specified, as we propose to do here, the above set up would be referred to as *Bayes empirical Bayes* [Deely and Lindley (1981)].

In this paper, we shall assume that the $\underset{\sim}{\theta}_t$'s are generated by a multivariate normal distribution with an unknown mean vector $\underset{\sim}{\lambda}=(\lambda_1,\ldots,\lambda_p)'$ and a known p x p covariance matrix $\underset{\sim}{V}$. Our uncertainty about $\underset{\sim}{\lambda}$ will also be described by a multivariate normal distribution with a mean vector $\underset{\sim o}{m}$ and covariance matrix $\underset{\sim o}{s}$. Both $\underset{\sim o}{m}$ and $\underset{\sim o}{s}$ have to be specified initially; however upon the receipt of data they will be updated according to Bayes law. Thus to summarize, a proper Bayesian description of the nonhomogeneous autoregressive process considered by us, goes as follows:

$$
\begin{aligned}
y_t &= \underset{\sim t}{\theta}' \underset{\sim t-1}{y}^{(p)} + u_t \text{ , with} \\
u_t &\sim N(0,\sigma_u^2) \quad \text{, where } \sigma_u^2 \text{ is specified;} \\
\underset{\sim t}{\theta} &\sim N(\underset{\sim}{\lambda},\underset{\sim}{V}) \quad \text{, where } \underset{\sim}{V} \text{ is specified, and} \\
\underset{\sim}{\lambda} &\sim N(\underset{\sim o}{m},\underset{\sim o}{s}) \quad \text{, where } \underset{\sim o}{m} \text{ and } \underset{\sim o}{s} \text{ are also specified.}
\end{aligned}
\tag{3.1}
$$

The above set-up can also be expressed as a *dynamic linear model* in the sense of Harrison and Stevens (1976) and therefore the Kalman Filter solution can be used for inference about $\underset{\sim}{\theta}_t$ given $y_1,\ldots,y_t$. To see this, we first rewrite (3.1) as

$$\begin{aligned} y_t &= \underset{\sim}{\theta}_t' \, \underset{\sim}{y}_{t-1}^{(p)} + u_t, \text{ with } u_t \sim N(0,\sigma_u^2), \text{ and} \\ \underset{\sim}{\theta}_t &= \underset{\sim}{\lambda} + \underset{\sim}{w}_t, \text{ with } \underset{\sim}{w}_t \sim N(\underset{\sim}{0},\underset{\sim}{V}), \end{aligned} \tag{3.2}$$

where the u_t's are independent of the $\underset{\sim}{w}_t$'s, $\underset{\sim}{\lambda}$ is independent of $\underset{\sim}{w}_t$ and $\underset{\sim}{\lambda} \sim N(\underset{\sim}{m}_o,\underset{\sim}{s}_o)$.

To cast (3.1) into the format of a Kalman Filter model, we let

$$\underset{\sim}{F}_t = (\underset{\sim}{y}_{t-1}^{(p)\prime},\underset{\sim}{0}'), \quad \bar{\underset{\sim}{\theta}}_t = \begin{pmatrix} \underset{\sim}{\theta}_t \\ \underset{\sim}{\lambda} \end{pmatrix},$$

$$\underset{\sim}{G}_t = \begin{pmatrix} \underset{\sim}{0} & \underset{\sim}{I}_p \\ \underset{\sim}{0} & \underset{\sim}{I}_p \end{pmatrix}, \quad \bar{\underset{\sim}{w}}_t = \begin{pmatrix} \underset{\sim}{w}_t \\ \underset{\sim}{0} \end{pmatrix},$$

$$\underset{\sim}{W}_t = \begin{pmatrix} \underset{\sim}{V} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix},$$

where $\underset{\sim}{I}_p$ is the p x p identity matrix and $\underset{\sim}{W}_t$ is the covariance matrix of $\bar{\underset{\sim}{w}}_t$, and we rewrite (3.2) as

$$\begin{aligned} y_t &= \underset{\sim}{F}_t \, \bar{\underset{\sim}{\theta}}_t + u_t \\ \bar{\underset{\sim}{\theta}}_t &= \underset{\sim}{G}_t \, \bar{\underset{\sim}{\theta}}_{t-1} + \bar{\underset{\sim}{w}}_t. \end{aligned} \tag{3.3}$$

Using the well known solution to the standard Kalman Filter model [see for example Meinhold and Singpurwalla (1983)], we have the result

that given $y(t) = (y_1,\ldots,y_t)$ and $y_o,\ldots,y_{-(p-1)}$

$$(\underset{\sim}{\theta}_t | y(t)) \sim N(\hat{\underset{\sim}{\theta}}_t, \underset{\sim}{\Sigma}_t), \text{ where}$$

$$\hat{\underset{\sim}{\theta}}_t = \underset{\sim}{m}_{t-1} + \frac{(\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\underset{\sim}{y}^{(p)}_{t-1}}{\sigma_u^2 + \underset{\sim}{y}^{(p)\prime}_{t-1} (\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\underset{\sim}{y}^{(p)}_{t-1}} (y_t - \underset{\sim}{m}'_{t-1}\, \underset{\sim}{y}^{(p)}_{t-1}) \tag{3.4}$$

and

$$\underset{\sim}{\Sigma}_t = (\underset{\sim}{s}_{t-1} + \underset{\sim}{V}) - \frac{(\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\underset{\sim}{y}^{(p)}_{t-1}\, \underset{\sim}{y}^{(p)\prime}_{t-1}\, (\underset{\sim}{s}_{t-1} + \underset{\sim}{V})}{\sigma_u^2 + \underset{\sim}{y}^{(p)\prime}_{t-1} (\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\underset{\sim}{y}^{(p)}_{t-1}}, \text{ with} \tag{3.5}$$

$$\underset{\sim}{m}_t = \underset{\sim}{m}_{t-1} + \frac{\underset{\sim}{s}_{t-1}\, \underset{\sim}{y}^{(p)}_{t-1}}{\sigma_u^2 + \underset{\sim}{y}^{(p)\prime}_{t-1} (\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\underset{\sim}{y}^{(p)}_{t-1}} (y_t - \underset{\sim}{m}'_{t-1}\, \underset{\sim}{y}^{(p)}_{t-1}), \text{ and} \tag{3.6}$$

$$\underset{\sim}{s}_t = \underset{\sim}{s}_{t-1} - \frac{\underset{\sim}{s}_{t-1}\, \underset{\sim}{y}^{(p)}_{t-1}\, \underset{\sim}{y}^{(p)\prime}_{t-1}\, \underset{\sim}{s}_{t-1}}{\sigma_u^2 + \underset{\sim}{y}^{(p)\prime}_{t-1} (\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\underset{\sim}{y}^{(p)}_{t-1}} . \tag{3.7}$$

Furthermore, the posterior distribution of $\underset{\sim}{\lambda}$ given $y(t)$ is

$$(\underset{\sim}{\lambda} | y(t)) \sim N(\underset{\sim}{m}_t, \underset{\sim}{s}_t)$$

where the updating formulas for $\underset{\sim}{m}_t$ and $\underset{\sim}{s}_t$ are given above. The covariance of $\underset{\sim}{\theta}_t$ and $\underset{\sim}{\lambda}$ given $y(t)$ is given by

$$\text{cov}(\underset{\sim}{\theta}_t, \underset{\sim}{\lambda} | y(t)) = \underset{\sim}{s}_{t-1} - \frac{(\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\, \underset{\sim}{y}^{(p)}_{t-1}\, \underset{\sim}{y}^{(p)\prime}_{t-1}\, \underset{\sim}{s}_{t-1}}{\sigma_u^2 + \underset{\sim}{y}^{(p)\prime}_{t-1} (\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\underset{\sim}{y}^{(p)}_{t-1}} . \tag{3.8}$$

The predictive density of y_{t+1} given $y(t)$ is of the form

$$(y_{t+1} | y(t)) \sim N[\underset{\sim}{m}'_t \underset{\sim}{y}^{(p)}_t,\ \underset{\sim}{y}^{(p)\prime}_t (\underset{\sim}{s}_t + \underset{\sim}{V})\, \underset{\sim}{y}^{(p)}_t + \sigma_u^2].$$

Under the assumption of a quadratic loss, $\hat{\underset{\sim}{\theta}}_t$ and $\underset{\sim}{m}_t$ are the Bayes estimators of $\underset{\sim}{\theta}_t$ and $\underset{\sim}{\lambda}$. When p=1, the Bayes estimator of $\underset{\sim}{\theta}_t$ simplifies to

$$\hat{\theta}_t = \pi_t \, m_{t-1} + (1-\pi_t) \, y_t/y_{t-1}, \text{ where} \tag{3.9}$$

$$\pi_t = \frac{\sigma_u^2}{\sigma_u^2 + (\sigma_\nu^2 + s_{t-1})y_{t-1}^2}, \text{ and}$$

σ_ν^2 is the variance of $\underset{\sim}{w}_t$ in (3.2).

Thus for a first order nonhomogeneous autoregressive process the Bayes estimator at time t is a weighted average of the prior mean of θ_t (namely m_{t-1}) and the least squares estimate y_t/y_{t-1}. The weight π_t is a function of the variance components σ_u^2, σ_ν^2 and s_{t-1}. If σ_u^2 gets small or $(\sigma_\nu^2 + s_{t-1})$ gets large, then π_t gets small and in (3.9) more weight is given to the least squares estimator. We also note that the Bayes estimator at time t is based on all the available data at time t, whereas the least squares estimator is based on y_t and y_{t-1} only. As a final comment, we note that the Bayes estimator $\hat{\underset{\sim}{\theta}}_t$ can be obtained for any order p of the process, irrespective of the value of N.

For the p-th order nonhomogeneous process with m=1 and N > 1, we assume, following Anderson (1978), that coefficients $\underset{\sim}{\theta}_t$ are identical for all cross-sectional units and write the model as

$$\underset{\sim}{y}_t = \underset{\sim}{Y}_{t-1}^{(N\times p)} \, \underset{\sim}{\theta}_t + \underset{\sim}{\bar{u}}_t, \tag{3.10}$$

where $\underset{\sim}{Y}_{t-1}^{(N\times p)} = (\underset{\sim}{y}_{t-1} \, \underset{\sim}{y}_{t-2} \cdots \underset{\sim}{y}_{t-p})$ is a N×p matrix and $\underset{\sim}{y}_t = (y_{1t} \, y_{2t} \cdots y_{Nt})'$.

The N-dimensional vector $\underset{\sim}{\bar{u}}_t = (\bar{u}_{1t}\ \bar{u}_{2t} \ldots \bar{u}_{Nt})'$ is assumed to be normally distributed with mean vector $\underset{\sim}{0}$ and a specified variance-covariance matrix, say $\underset{\sim}{U}$.

By judging $\{\underset{\sim}{\theta}_t\}$ as an exchangeable sequence, by replacing y_t by the N-dimensional vector $\underset{\sim}{y}_t$, $\underset{\sim}{y}_{t-1}^{(p)'}$ by the (N×p) matrix $\underset{\sim}{Y}_{t-1}^{(N\times p)}$ and u_t by $\underset{\sim}{\bar{u}}_t$ in (3.2), we can cast the above model into the framework of the Kalman filter. We then appeal to the Kalman filter solution, and obtain the posterior distribution of $\underset{\sim}{\theta}_t$ given $\underset{\sim}{y}(t) = (\underset{\sim}{y}_1, \underset{\sim}{y}_2, \ldots, \underset{\sim}{y}_t)$ as a normal with mean

$$\hat{\underset{\sim}{\theta}}_t = \underset{\sim}{m}_{t-1} + (\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\ \underset{\sim}{Y}_{t-1}^{(N\times p)'}\ \underset{\sim}{P}_t^{-1}\ (\underset{\sim}{y}_t - \underset{\sim}{Y}_{t-1}^{(N\times p)}\ \underset{\sim}{m}_{t-1}) \qquad (3.11)$$

and variance

$$\underset{\sim}{\Sigma}_t = (\underset{\sim}{s}_{t-1} + \underset{\sim}{V}) - (\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\ \underset{\sim}{Y}_{t-1}^{(N\times p)'}\ \underset{\sim}{P}_t^{-1} \underset{\sim}{Y}_{t-1}^{(N\times p)} (\underset{\sim}{s}_{t-1} + \underset{\sim}{V}), \qquad (3.12)$$

where

$$\underset{\sim}{P}_t = \underset{\sim}{Y}_{t-1}^{(N\times p)}\ (\underset{\sim}{s}_{t-1} + \underset{\sim}{V})\ \underset{\sim}{Y}_{t-1}^{(N\times p)'} + \underset{\sim}{U}\ ,$$

$$\underset{\sim}{m}_t = \underset{\sim}{m}_{t-1} + \underset{\sim}{s}_{t-1}\ \underset{\sim}{Y}_{t-1}^{(N\times p)'}\ \underset{\sim}{P}_t^{-1}\ (\underset{\sim}{y}_t - \underset{\sim}{Y}_{t-1}^{(N\times p)} \underset{\sim}{m}_{t-1}),\ \text{and}$$

$$\underset{\sim}{s}_t = \underset{\sim}{s}_{t-1} - \underset{\sim}{s}_{t-1}\ \underset{\sim}{Y}_{t-1}^{(N\times p)'}\ \underset{\sim}{P}_t^{-1}\ \underset{\sim}{Y}_{t-1}^{(N\times p)}\ \underset{\sim}{s}_{t-1}\ .$$

We note that $\underset{\sim}{m}_t$ and $\underset{\sim}{s}_t$ are the posterior mean and covariance matrix of $\underset{\sim}{\lambda}$.

For p=1 and $\underset{\sim}{U} = \sigma_u^2\ \underset{\sim}{I}_N$, $\hat{\underset{\sim}{\theta}}_t$ simplifies to

$$\hat{\theta}_t = \pi_t\ m_{t-1} + (1 - \pi_t)\ \frac{\underset{\sim}{y}_{t-1}'\ \underset{\sim}{y}_t}{\underset{\sim}{y}_{t-1}'\ \underset{\sim}{y}_{t-1}}\ ,$$

where

$$\pi_t = \frac{\sigma_u^2}{\sigma_u^2 + \underset{\sim}{y}'_{t-1}\,\underset{\sim}{y}_{t-1}\,(s_{t-1} + \sigma_v^2)}\,, \text{ and}$$

$\dfrac{\underset{\sim}{y}'_{t-1}\,\underset{\sim}{y}_t}{\underset{\sim}{y}'_{t-1}\,\underset{\sim}{y}_{t-1}}$, is the least-squares estimate of $\underset{\sim}{\theta}_t$.

Thus the Kalman filter solution results in a shrinkage estimator which is a linear combination of the prior mean and the least squares estimator.

4. Bayesian Estimation in Nonhomogeneous Autoregressive Processes Assuming Autoregression of Coefficients.

Consider the first order nonhomogeneous autoregression process (2.1) with m=N=1 and assume a time pattern to the $\underset{\sim}{\theta}_t$'s, where now $\underset{\sim}{\theta}_t = \theta_{1t} = \theta_t$. Specifically, let $\theta_t = \alpha\theta_{t-1} + w_t$, where α is unknown and the innovation w_t is normal with mean 0 and known variance σ_w^2. Thus the model considered here is of the form

$$y_t = \theta_t y_{t-1} + u_t$$

and (4.1)

$$\theta_t = \alpha\theta_{t-1} + w_t,$$

where $u_t \sim N(0,\sigma_u^2)$, $w_t \sim N(0,\sigma_w^2)$ with σ_u^2 and σ_w^2 known and y_o is known. The sequences $\{u_t\}$ and $\{w_t\}$ are assumed independent. Uncertainty about θ_0 is described by a normal density with mean $\hat{\theta}_0$ and variance $\hat{\Sigma}_0$ which are both specified.

The above set up is that of a Kalman filter model except for the fact that α is unknown. Suppose that our uncertainty about α is described by $p(\alpha)$ a prior distribution for α given some background information. Then given some data $y(t)$, where we recall that $y(t) = (y_1,\ldots,y_t)$, our goal is to make inferences about θ_t and the future observations y_{t+1}, $y_{t+2},\ldots,$. Extending consideration to α, the posterior distribution of θ_t is

$$p(\theta_t|y(t)) = \int p(\theta_t|y(t),\alpha)\ p(\alpha|y(t))d\alpha, \tag{4.2}$$

where $p(\theta_t|y(t),\alpha)$ is obtained by the usual Kalman filter solution with α assumed known and $p(\alpha|y(t))$, the posterior distribution of α given $y(t)$, is obtained via Bayes law as

$$p(\alpha|y(t)) \propto L(\alpha;\ y(t))p(\alpha), \tag{4.3}$$

with $L(\alpha;\ y(t))$ as the likelihood function of α. For the ordinary Kalman filter with α known, the predictive distribution of y_i given $y(i-1)$ is

$$p(y_i|y(i-1),\alpha) = \int_{\theta_i} p(y_i|\theta_i,y(i-1),\alpha)p(\theta_i|y(i-1),\alpha)d\theta_i\ , \tag{4.4}$$

where $(\theta_i|y(i-1),\alpha) \sim N(\alpha\hat{\theta}_{i-1},R_i)$ where $R_i = \alpha^2\ \hat{\Sigma}_{i-1} + \sigma_w^2$; the quantities $\hat{\theta}_{i-1}$ and $\hat{\Sigma}_{i-1}$ are obtained recursively via the relationship $(\theta_{i-1}|y(i-1),\alpha) \sim N(\hat{\theta}_{i-1},\ \hat{\Sigma}_{i-1})$. Specifically,

$$\hat{\theta}_{i-1} = \frac{\alpha\hat{\theta}_{i-2}\ \sigma_u^2 + R_{i-1}\ y_{i-1}y_{i-2}}{y_{i-2}^2\ R_{i-1} + \sigma_u^2} \tag{4.5}$$

and

$$\hat{\Sigma}_{i-1} = \frac{R_{i-1}\sigma_u^2}{y_{i-2}^2 R_{i-1} + \sigma_u^2} . \tag{4.6}$$

It now follows from the above that

$$(y_i | y(i-1),\alpha) \sim N(\alpha\hat{\theta}_{i-1} y_{i-1}, y_{i-1}^2 R_i + \sigma_u^2) \tag{4.7}$$

and so the likelihood of α may be written

$$L(\alpha; y(t)) = \prod_{i=1}^{t} p(y_i | y(i-1),\alpha)$$

where the terms in the product are determined by (4.7).

We may now write (4.2) as

$$p(\theta_t | y(t)) = \frac{\int p(\theta_t | y(t),\alpha)\ L(\alpha;y(t))p(\alpha)d\alpha}{\int L(\alpha;\ y(t))\ p(\alpha)d\alpha} . \tag{4.8}$$

Any reasonable prior distribution of α that we may consider leads us to integrals in (4.8) which cannot be expressed in closed form. The same is also true when we consider the predictive distribution of y_{t+1} given y(t); that is, the ratio of the integrals

$$p(y_{t+1} | y(t)) = \frac{\int p(y_{t+1} | y(t),\alpha)\ L(\alpha;y(t))\ p(\alpha)d\alpha}{\int L(\alpha;\ y(t))\ p(\alpha)d\alpha} . \tag{4.9}$$

One way to deal with the evaluation of the integrals (4.8) and (4.9) is via numerical methods. Another way is via an approximation due to Lindley (1980) which performs well when $t \to \infty$. For convenience, we give below an overview of Lindley's approximation.

Lindley (1980) develops asymptotic expansions for the ratio of integrals that occur frequently in Bayesian analyses. He considers ratios of the form

$$\int w(\alpha)\ e^{L(\alpha)}d\alpha / \int p(\alpha)e^{L(\alpha)}d\alpha \tag{4.10}$$

where α is an (unknown) parameter and $L(\alpha)$ is the logarithm of its likelihood, with dependence on $y(t)$, the data, being suppressed; that is

$$L(\alpha) = \log L(\alpha;y(t)) = \sum_{i=1}^{t} \log \{p(y_i|y(i-1),\alpha)\}.$$

The quantity $w(\alpha) \underline{\underline{\text{def}}}\ p(\alpha)\ u(\alpha)$ and $u(\alpha)$ is some function of α that is of interest. For example, if $u(\alpha) = \alpha$, then (4.10) is the mean of the posterior distribution of α.

Lindley's approximation is concerned with the asymptotic behavior of (4.10) as $t \to \infty$. This is facilitated by the fact that asymptotically, $L(\alpha)$ concentrates around $\hat{\alpha}$, its maximum, assuming that the maximum is unique. The idea is to obtain a Taylor's series expansion of all the functions of α in (4.10). Let $H(\alpha) = \log p(\alpha)$; then (4.10) can also be written as

$$\int u(\alpha)e^{L(\alpha)+H(\alpha)}d\alpha / \int e^{L(\alpha)+H(\alpha)}d\alpha. \tag{4.11}$$

If we let $u(\alpha) = p(\theta_t|y(t),\alpha)$ [or $p(y_{t+1}|y(t),\alpha)$], then an approximation to (4.11) is

$$u(\hat{\alpha}) - \frac{u_2 + 2u_1H_1}{2L_2} + \frac{u_1L_3}{2L_2^2}, \tag{4.12}$$

where

$$u_i = \left.\frac{d^i u(\alpha)}{d\alpha^i}\right|_{\alpha=\hat{\alpha}} , \quad i=1,2$$

$$H_1 = \left.\frac{dH(\alpha)}{d\alpha}\right|_{\alpha=\hat{\alpha}} , \text{ and } L_i = \left.\frac{d^i L(\alpha)}{d\alpha^i}\right|_{\alpha=\hat{\alpha}} , \quad i=2,3.$$

A convenient prior for α is the uniform on [a,b]. In this case $H(\alpha)$ is a constant. When $u(\alpha) = E(\theta_t|y(t),\alpha)$, (4.12) gives us approximately $E(\theta_t|y(t))$, the *optimal adaptive Kalman filter estimate.* When $u(\alpha) = p(y_{t+1}|y(t),\alpha)$, (4.12) gives us approximately $p(y_{t+1}|y(t))$. The quantities $E(\theta_t|y(t),\alpha)$ and $p(y_{t+1}|y(t),\alpha)$ are given by (4.5) and (4.7) respectively. To obtain $E(y_{t+1}|y(t))$, the *predictive mean,* we set $u(\alpha) = E(y_{t+1}|y(t),\alpha)$.

REFERENCES

ANDERSON, T. W. (1978). Repeated Measurements on Autoregressive Processes. *Journal of the American Statistical Association, Vol. 73,* pp. 371-378.

BROEMELING, L. D. (1985). Bayesian Analysis of Linear Models. Marcel Dekker, Inc., New York.

DEELY, J. J. and D. V. LINDLEY (1981). Bayes Empirical Bayes. *Journal of the American Statistical Association, Vol. 76,* pp. 833-841.

DEMPSTER, A. P., N. M. LAIRD and D. G. RUBIN (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Ser. B, Vol. 39,* pp. 1-38.

HARRISON, P. J. and C. F. STEVENS (1976). Bayesian Forecasting (with discussion). *Journal of the Royal Statistical Society, Ser. B, Vol. 38,* pp. 205-247.

HORIGOME, M., N. D. SINGPURWALLA and R. SOYER (1985). A Bayes Empirical Bayes Approach for (Software) Reliability Growth. Computer Science and Statistics: Proceedings of the 16th Symposium on the Interface (L. Billard, Ed.)., pp. 47-56.

HSIAO, C. (1986). Analysis of Panel Data, in press.

LINDLEY, D. V. (1980). Approximate Bayesian Methods. *Trabajos Estadistica, Vol. 32,* pp. 223-237.

LIU, L. M. and G. C. TIAO (1980). Random Coefficient First Order Autoregressive Models. *Journal of Econometrics, Vol. 13,* pp. 305-325.

MAGILL, D. T. (1965). Optimal Adaptive Estimation of Sampled Stochastic Processes. *IEEE Transactions on Automatic Control, Vol. AC-10,* pp. 434-439.

MEHRA, R. K. (1972). Approaches to Adaptive Filtering. *IEEE Transactions on Automatic Control, Vol. AC-17,* pp. 693-698.

MEINHOLD, R. J. and N. D. SINGPURWALLA (1983). Understanding the Kalman Filter. *American Statistician, Vol. 37,* pp. 123-127.

MORRIS, C. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association, Vol. 78,* pp. 47-65.

SHUMWAY, R. H. (1983). Some Applications of EM Algorithm to Analyzing Incomplete Time Series Data. Symposium on Time Series Analysis of Irregularly Observed Data, Institute of Statistics, Texas A&M University.

Technical Reports

U.S. Army Research Office - Contracts DAAG29-82-K-0156 and DAAG29-82-K-0239

1. "Maximum Likelihood Estimators and Likelihood Ratio Criteria for Multivariate Elliptically Contoured Distributions," T. W. Anderson and Kai–Tai Fang, September 1982.

2. "A Review and Some Extensions of Takemura's Generalizations of Cochran's Theorem," George P.H. Styan, September 1982.

3. "Some Further Applications of Finite Difference Operators," Kai–Tai Fang, September 1982.

4. "Rank Additivity and Matrix Polynomials," George P.H. Styan and Akimichi Takemura, September 1982.

5. "The Problem of Selecting a Given Number of Representative Points in a Normal Population and a Generalized Mills' Ratio," Kai–Tai Fang and Shu–Dong He, October 1982.

6. "Tensor Analysis of ANOVA Decomposition," Akimichi Takemura, November 1982.

7. "A Statistical Approach to Zonal Polynomials," Akimichi Takemura, January 1983.

8. "Orthogonally Invariant Minimax Estimator of the Covariance Matrix of a Multivariate Normal Population," Akimichi Takemura, April 1983.

9. "An Orthogonally Invariant Minimax Estimator of the Covariance Matrix of a Multivariate Normal Population," Akimichi Takemura, April 1983.

10. "Relationships Among Classes of Spherical Matrix Distributions," Kai–Tai Fang and Han–Feng Chen, April 1984.

11. "A Generalization of Autocorrelation and Partial Autocorrelation Functions Useful for Identification of ARMA(p,q) Processes," Akimichi Takemura, May 1984.

12. "Methods and Applications of Time Series Analysis Part II: Linear Stochastic Models," T. W. Anderson and N. D. Singpurwalla, October 1984.

13. "Why Do Noninvertible Estimated Moving Averages Occur?" T. W. Anderson and Akimichi Takemura, November 1984.

14. "Invariant Tests and Likelihood Ratio Tests for Multivariate Elliptically Contoured Distributions," Huang Hsu, May 1985.

15. "Statistical Inferences in Cross–lagged Panel Studies," Lawrence S. Mayer, November 1985.

16. "Notes on the Extended Class of Stein Estimators," Suk–ki Hahn, July 1986.

17. "The Stationary Autoregressive Model," T. W. Anderson, July 1986.

18. "Bayesian Analyses of Nonhomogeneous Autoregressive Processes,," T. W. Anderson, Nozer D. Singpurwalla, and Refik Soyer, September 1986.